The Alberta Journal of Educational Research

Vol. IV *September, 1958*

THE COMMITTEE ON EDUCATIONAL RESEARCH
Faculty of Education
University of Alberta

ACKNOWLEDGMENT

This publication was made possible by funds granted by Carnegie Corporation of New York. That Corporation is not, however, the author, owner, publisher, or proprietor of this publication, and is not to be understood as approving by virtue of its grant any of the statements made or views expressed therein.

The Alberta Journal of Educational Research

THE COMMITTEE ON EDUCATIONAL RESEARCH

Faculty of Education
University of Alberta

VOL. IV, No. 3 SEPTEMBER, 1958

Editor: H. E. Smith

CONTENTS

PAGE

Editorial 129
G. M. Dunlop

Organization for Educational Research in Universities of Midwestern United States 131
R. S. MacArthur

The Development of the English Program in the Secondary Schools of Alberta 142
Stanley W. Sawicki

The Role of the Provincially Appointed Superintendent of Schools in Larger Units of Administration in Canada 152
C. P. Collins

The Wandering Quotients Settle Down 158
C. C. Anderson

The Effect of Grouping on Variability in Achievement at the Grade III Level 162
S. C. T. Clarke

A Note on the Use in Alberta of the Sequential Tests of Educational Progress: Essay Test 172
Donald B. Black

Book Review 181
W. E. Hodgson

Editorial

G. M. Dunlop

The Teacher and the Research Program

This is an anniversary in the history of educational research in Alberta. It marks the end of the fifth year since the whole idea of an Alberta research organization was first discussed on an informal basis by representatives of what later became the five sponsoring groups: Department of Education, The Alberta Trustees' Association, The Alberta Teachers' Association, The Alberta Federation of Home and School Associations, and the University of Alberta. The five year period has witnessed memorable events—the completion of the organization, the enrichment of the research program, the launching of the *Alberta Journal of Educational Research, the Alberta Newsletter on Educational Research,* and now, the *Alberta Monographs on Educational Research,* the Carnegie award, the establishment of the Carnegie research fellowships, and, this autumn, the introduction of the basic machines for card punching and sorting, which facilitate the speedy and accurate analysis of educational data. Taken individually or collectively, these occurrences represent rapid and marked progress in the development of the research program, and augur well for its future.

What is the relationship between the research program and the classroom teacher? This is a rhetorical question which invites a restatement of the credo of the teacher in so far as it concerns research.

1. Teachers are the *ultimate evaluators* and *consumers of research.* Today's classroom practice is based on the research of yesterday; tomorrow's advances will reflect research insights of today.
2. Educational research *becomes possible* only through the *consent* and *cooperation* of the teachers.
3. The *climate of opinion* of a school system and its administration toward research is a direct outcome of the professional research mindedness of the teaching staffs.
4. Much *financial support of research* comes directly from the *contributions* of the *teachers' central* and *local organizations,* and indirectly from the contributions of organizations and industries influenced by teacher opinion.
5. The *nature* and *direction* of the on-going research program is influenced continuously by the teachers' need for answers to problems for which suitable answers simply do not exist.

In these contentious days an occupational hazard of our profession is the temptation to lose faith, however briefly, in certain of our educational practices, as they come under the attention of our critics. It is the mark of a true profession that practice is based upon the accumulated facts, knowledge and insight which experience and research have provided, and that practice must be under the continuous appraisal of research and amenable to change based upon its findings.

The conclusion that there has been too little educational research of the requisite high quality is forced upon all students of education. From press reports it would seem that the Cameron Commission has again and again encountered questions for which they could find no definitive answers. The reason—too little research, and that of a haphazard, hit and miss kind which does not stay with a problem enough to reach the final answers.

Thanks to the splendid spirit of cooperation found in this province, we have made an excellent beginning in educational research. The teachers and their professional association have been in the forefront of the movement from the start. The next five years can witness the coming to full maturity and full fruition of the research movement. You can assist in the realization of this objective by a continuance and even an increase in your support, whether through active participation in research, cooperation with on-going projects, development of a favorable climate of opinion for research, or in increased financial support from your locals, your provincial organization, or from agencies in society concerned with the welfare of education.

ORGANIZATION FOR EDUCATIONAL RESEARCH IN UNIVERSITIES OF MIDWESTERN UNITED STATES

R. S. MacArthur
Faculty of Education
University of Alberta

Introduction: A Problem

"Fundamentally, individuals do research." "Research (in a university) is a function of the individual professor or of groups of professors." These two statements were recently made by Deans of Education in two large American universities. They underline a viewpoint held by many persons — that productive research (in a university setting at any rate) depends basically upon the creativity and imagination of individuals, or of small groups of individuals in close communication.

On the other hand, S. C. T. Clarke (1) has indicated four main trends in current educational research—trends toward specialization, complexity, magnitude, and dispersion.

If the foregoing observations are correct, and this writer believes that they are, therein lies a problem of considerable proportions in the conduct of educational research today. How can modern educational research—specialized, complex, big, and dispersed—be so organized as to provide optimum scope for the creative and imaginative minds of curious and inquiring individuals? This article seeks to describe the approaches toward solution of this problem being made by ten universities in midwestern United States.

In the spring of 1958 the writer was awarded a Travel Grant by the Carnegie Corporation of New York, ,for the purpose of studying current trends in educational research in midwestern United States, with special reference to the area of educational psychology. During the two-and-one-half months trip, visits of approximately one week each were made to the following institutions: University of Minnesota, University of Wisconsin, University of Chicago, Purdue University, Ohio State University, University of Pittsburgh, University of Michigan, Science Research Associates, and Evanston (Illinois) Public Schools. The study was conducted through two main avenues: (a) interviews with persons selected under the guidance of the respective Deans of Education, and (b) perusal of documents provided or recommended by these persons. Approximately two hundred people were interviewed, in sessions ranging from a half-hour to two hours, and on-the-spot notes were prepared during each interview. A four-foot file of documents was amassed during the trip;

most of those pertaining to any particular institution were skimmed while the writer was still at that institution, and in addition they provided a wealth of detail for later reference. Supplementary avenues of study included attendance at conferences, participation in seminars and meetings, visits to on-going research projects, conversations with graduate students, and so forth. Research activities proceeding through Departments of Psychology and Graduate Schools, and having relevance to Education, were of special interest to the writer.

Three main kinds of questions were addressed to interviewees, with emphasis depending upon the interests of the interviewee: (a) questions concerning organization for research, (b) questions concerning specific research projects, and ranging from theoretical backgrounds and design through to findings, (c) questions concerning preparation programs for research workers.

The writer would find little with which to disagree in Dr. Clarke's aforementioned analytical discussion of trends and problems in current American educational research. The present paper, at a more descriptive level, reports this writer's impressions with respect to various practices in organization for educational research in the ten universities visited, making certain generalizations concerning both administrative structures and other facilitating provisions, generalizations which might or might not always apply to specific universities. Observations and comments are restricted to the university picture, and are not intended to apply necessarily to research conducted by school systems, governmental agencies, or other institutions.

Key Role of Individual Interests

Reference has already been made to the viewpoints of two Deans of Education with regard to the unique importance of the individual professor in educational research in universities. This stress upon the necessity of having, as the "principal investigator" in a project, a person who is in close touch with both the theoretical bockgrounds of his project and the day-to-day operation of his project was common throughout the universities visited. J. B. Stroud (Iowa), for example, felt that generally speaking the most qualified man to do research in a particular field is the man who is studying and teaching in that field, and not a man set aside in a research bureau. On the operational side, Dale Harris (Minnesota) thought that a person administering a research project must himself get working with the data, beside the assistants and clerks. If he is merely administering, the research is liable to follow routine and obvious directions; insights into the new leads and promising hypotheses are much more likely when the researcher is closely acquainted with the emerging data of his project.

William McLure, Director of the University of Illinois Bureau of Educational Research, was somewhat sceptical of massive organizations for educational research, and would gamble on relatively small projects with a very few professional people responsible for each. He felt that outside money for research should be pin-pointed to individuals, and that a research director must match financial support with readiness of individuals to do research. The work of the Illinois Bureau is almost entirely individual in nature, with very few projects extending beyond the active involvement of three or four professional persons. Even in the Institute for Social Research at the University of Michigan, where the thirty to forty key people are full-time career employees of the Institute with their main commitment definitely to research within the Institute structure, the effective unit of organization is the project, with each project placed in the charge of one person whose career interests and scholarship have usually been in line with the topic of investigation, and who has played an active part in bringing about the grant and contract for that particular project.

Regardless of the administrative structures evolved to further educational research, then, it appeared to be definite policy in each university to foster a link between the professional interests of individual staff members on the one hand, and on the other hand projects of such size and nature that the staff member could maintain active involvement in the research. The main administrative structures, employed in various combinations in the several universities, will be outlined in the following sections.

One Professor with a Group of Graduate Students

This was, of course, the most common form of organization, occurring in all universities. Practices differed considerably however, even within the same graduate school, in the degree to which the thesis investigations of the students were related to specific projects being conducted by the professor. At one extreme, graduate students working as research assistants on projects directed by the professor might be required to keep their work as research assistants distinctly separate from their thesis investigations. At the other extreme, and more commonly, being accepted as a research assistant to a professor might be synonymous with undertaking a thesis investigation as part of one of the research projects of that professor, whether or not the student was initially interested in the topic. Usually, however, the professor had definite and long-term research interests; he obtained funds for research assistants to help him in pursuing these interests; most of these research assistants would be graduate students who later found their thesis investigations directly in one of his projects but others would not.

When asked about the degree to which he thought that graduate student thesis investigations should be patterned, or even dictated, by staff interests, John Darley, Associate Dean of the Graduate School at Minnesota, replied that he thought this a somewhat unimportant issue. He felt that the main point in this connection was to maintain high research productivity and morale on the part of the faculty, and that the question could best be left to the discretion of individual faculty members. Moody Prior, Dean of the Graduate School at Northwestern, also felt that the degree to which the graduate student might be "used as a tool" by the professor depended largely upon the attitude of the professor, ,but could be controlled somewhat by limiting the number of graduate dissertations the professor was allowed to supervise.

Bureaus of Educational Research

Superficially one might expect that if a school or college of education wished to foster educational research as one of its major functions, an obvious step would be to set up a bureau of educational research, whose major purpose would be to conduct research. Yet one Dean of Education stated flatly that in his opinion a bureau of educational research puts a straight-jacket on research, while another Dean of Education remarked that many people did not feel particularly favorable toward such bureaus. In only three of the ten universities visited were there effectively functioning bureaus of educational research.

The University of Illinois Bureau of Educational Research usually has a regular staff of ten academic members of professorial rank, eight part-time research assistants, and twenty-one full-time or part-time clerical persons. This regular staff is usually augmented, on a temporary basis, by about two professors, eight research assistants, and twelve clerical assistants. Each member of the professorial staff of the Bureau usually teaches one course per semester. The research activities of the Bureau may be described as an aggregate of long-term research programs of the individual professors, assisted by graduate students and other staff assigned to them. The regular University budget of the Bureau is considerably supplemented by funds from outside the University. Non- Bureau members of the College of Education may secure from the Bureau financial help for research projects on the same basis as many Bureau members.

The staff of the University of Minnesota Bureau of Educational Research includes the Director, usually about nine research fellows and part-time research assistants, an office supervisor, and clerical staff as needed. The Ohio State University Bureau of Educational Research and Service, a large organization with twenty-six senior staff, has been re-organized into twelve divisions. Of late years this

Bureau has become mainly a service bureau, and the recent reorganization is aimed toward bringing back research as one of its major functions. The University of Iowa is in the process of closing out its Bureau of Educational Research, the University of Michigan had such a Bureau at one time but closed it, while the University of Wisconsin is in the process of launching what it calls Cooperative Research and Services.

In some institutions, then, bureaus of educational research seem to be performing valuable but varied services; other institutions consider that educational research can be better served without such bureaus. Where a bureau is in effective operation, it is but one of several organizations extensively engaged in educational research in the institution. The main services offered by most of the bureaus are included in the following: (1) Definite allocation of the time of certain staff members and graduate students for research, (2) Provision of space for housing research activities, (3) Provision for communication amongst research workers (seminars, newsletters, bulletins), (4) Training for research workers, (5) Housing of reference libraries containing journals, specimen tests and other materials used primarily by research workers, (6) Provision of help for non-bureau members in designing research projects and proposals, (7) Provision of financial help for bureau and non-bureau members, (8) Taking over certain actual operational aspects of research projects for non-bureau members, e.g. distribution, collection and tabulation of questionnaires, or certain routine calculations and analyses, (9) Housing and operation of certain equipment (desk calculators, inexpensive IBM equipment), (10) Provision of a central office to which major research problems from the field may be referred.

On-Going Organizations with More Specific Interests

Certainly one of the most productive of the administrative arrangements for fostering research related to education, as observed by the writer, was through certain long-term organizations, working in the university setting, but usually having interests more specific than educational research in general. In many instances research was considered as one of several main functions. A number of these organizations which were visited by the writer are listed below. Their names will indicate something of the decidedly varied nature of their interests.

One group of such organizations was concerned with research directly related to elementary and/or secondary education. This group included: at Minnesota—the Laboratory Schools; at Chicago —the Teaching-Learning Laboratory, the Reading Clinic, the Mid-West Administration Center, and the School Improvement Program;

at Iowa—the Testing Program, and the Iowa Reading Laboratory; at Illinois—the University High School, and the Illinois Curriculum Project; at Ohio State—the Center for Educational Administration, and the University School; at Michigan—the University Elementary School.

For a second group, their educational research activities seemed to spring initially from a sort of action research into problems of higher education. This group included: at Minnesota—the Bureau of Institutional Research, the Student Personnel Office, and the Student Counseling Bureau; at Chicago—the Examiner's Office; at Purdue—the Division of Educational Reference; at Ohio State—the University Counseling and Testing Service; at Michigan—the Bureau of Psychological Services.

A third group of such organizations, some of them very large and carrying on extensive research programs, either were engaged in research centred in other disciplines but having rather immediate relevance for education, or included educational research in their research programs. This group included: at Minnesota—the Institute of Child Development and Welfare, and the Department for the Study of Exceptional Children; at Chicago—the Committee on Human Development, and the Orthogenic School; at Northwestern—the School of Education-Department of Psychology Carnegie Project; at Iowa—the Child Welfare Research Station; at Illinois—the Institute for Research on Exceptional Children, and the Institute for Personality and Ability Testing; at Purdue—the Department of Child Development and Family Life; at Pittsburg—the American Institute for Research; at Michigan—the Institute for Social Research, the The Mental Health Research Institute, and the Social Sciences Research Project.

With such obviously varied interests, it is apparent that the internal administrative structures of these organizations, and their relationships with their respective universities, would vary tremendously. But each appeared to have rather definite interests which provided focus and depth for their research programs. It was in these contexts that large-scale, long-term, multi-disciplinary team projects with numbers of smaller projects co-ordinated around them, were most apparent.

Ad Hoc Committees for Specific Projects

With the stress that universities seem to place upon the stimulation of individual interests, it was not surprising to find that the favorite form of administrative structure for research, beyond the single professor with his graduate students, consisted of the temporary banding together of two or three or four professors for the purpose of carrying out a specific project in which they found a com-

mon interest. These ad hoc committees, because of their temporary nature, could and did cut across any more permanent organization. Usually one or two professors were named as the principal investigators, with the others playing supporting roles.

Such committees could soon become inbred and cliquish, and their successful operation over a period of time depended upon easy communication amongst staff members, both across departments and across schools or colleges, together with reasonable freedom from time pressures.

Standing Committees for Educational Research

Standing committees are of course a time-honored administrative device. Three of the institutions visited were experimenting somewhat gingerly with collegewide standing committees for educational research. In each instance the committees had been in operation for approximately a year or less.

In the University of Illinois College of Education,four co-ordinating positions, each with an advisory committee, have recently been set up: one for graduate programs in education, one for undergraduate programs in education, one for field service activities, ,and one for research activities. At Northwestern and at Michigan educational research activities had been loosely co-ordinated through the Dean of Education, but recently established research committees were attempting to define their roles in this connection. Thus, in each instance, the standing committees on educational research seemed to be functioning as extensions of the co-ordinating function of the Dean, with particular reference to research.

While the duties of these committees were as yet by no means clear, the committees appeared to be aimed toward gathering information and clarifying college policy with respect to the following: (1) Research programs and projects—what research is currently going on, or has recently been completed in the college? what projects are being planned for the future, or would be planned if facilities and support were available? on what major programs should the college concentrate in the near future, while at the same time encouraging certain more specific and isolated projects? (2) Financial support—what funds from where are now being used by whom for what projects? what major available sources of financial support are not being fully tapped? what policies should be adopted with respect to budgets for research? how can staff be best advised concerning obtaining support for specific projects? (3) Research assistants—what are present practices with respect to research assistants? where are research assistants most needed, and how many are needed? how can research assistants be used so as to provide them with the best type of training? (4) Communication—how can staff

members be kept informed of grants received or available, and other research developments in the college? how can the college be kept informed about major research programs, new sources of funds, and other major research developments across the country? (5) Organization—what administrative structures can best facilitate research in the college? (6) Motivation for research—each committee member was expected to be himself actively engaged in conducting research.

Co-ordinators of Educational Research

Experimentation with a position called Co-ordinator of Educational Research was another recent administrative development, with its aim rather similar to that of the standing committees, namely, to increase the efficiency of the research program of the college while at the same time encouraging the research interests of individual professors and other organizations within the college.

At the University of Illinois the Co-ordinator of Research in the College of Education was operating in conjunction with an advisory committee, as indicated in the previous section. The University of Michigan Committee on Education Research was considering creating the position of Co-ordinator of Educational Research, which esentially would be that of executive officer of the Committee. Several other institutions, e.g. Purdue University, either had or were planning to have a person functioning as co-ordinator of educational research, with part of his time and responsibility definitely allocated for this work, although he was not officially given any title.

The College of Education in Ohio State University has nine Departments; in a recent re-organization of the College, using a sort of line-and-staff arrangement, there are now six staff positions: an Associate Dean, a Co-ordinator of the Instructional Program, a Co-ordinator of Student Field Experience, a Co-ordinator of Student Personnel, a Co-ordinator of Research, and a Co-ordinator of Field Studies and Services. Each of these staff positions is assigned a part of the leadership function of the Dean's Office. Each is concerned both with development and with operation. Thus Ross Mooney, the Co-ordinator of Research, is concerned with development and promotion of research throughout the College, and with those operational matters which are inter-departmental in nature.

As with the standing committees, so with the co-ordinators of research, their roles are still in an early evolutionary stage, but seem to be mainly those of: (1) Stimulating and encouraging research by others in the college, (2) Advising with respect to research design, (3) Assisting in the preparation of proposals for financial assistance, (4) Promoting research grants from outside agencies, (5) Handling

administrative interna with respect to research, (6) Teaching (about three hours a week) such subjects as Statistics or Testing and Measurement, (7) Personnally conducting some research, though not directing large projects. It is clearly the intention that the function of this position should be that of "co-ordinating," in the dictionary sense of "bringing parts into proper relation," rather than that of "directing," in the dictionary sense of "ordering to be done."

Consultants on Design

It was common practice in the universities visited (and particularly evident at Minnesota, Wisconsin, Chicago, Iowa, Illinois, and Ohio State), to have one or more professors whose definite responsibility it was to act, when their advice was sought, in the capacity of consultants on research design, and whose work-load was so adjusted that they could carry out this function. These professors assisted in the planning phases of most research projects: both staff projects and student thesis investigations, which were carried on in the college. They usually taught such subjects as Statistics or Testing and Measurement from three to six hours a week, served as consultants on design, and conducted certain research projects of their own.

Other Facilitating Provisions

Given suitable administrative structures, certain other facilitating provisions are almost necessary but certainly not sufficient for the conduct of good research. These include time, research assistants, space, equipment, and finance. Perhaps the most important of these is time for research. The University of Wisconsin Department of Education has adopted a policy for determining the work-load of Professional staff based upon a full load of fifteen work-units per week, this load being composed of the basic elements of Teaching, Research, and Field Service. The net result of this policy is a normal classroom teaching load of nine hours per week for staff members doing graduate work but not engaged in research, and a classroom teaching load of three to six hours per week for those engaged in research. A similar pattern of teaching commitment was common throughout the universities visited: for staff members engaged to any appreciable extent in research, their classroom teaching load was three to six hours per week. With respect to yearly commitments the following policy was common throughout: the staff member was engaged at an annual salary paid in twelve equal monthly instalments; for this annual salary he was committed to nine months of work for the university; for approved research done during the remaining three months, he was paid (from outside grants or from university funds) at the rate of one-ninth of his annual salary for each month of work.

In most of the schools or colleges of education visited (though not in all) it was usual for each professor engaged in research to have from two to six research assistants working for him. The standard rate of pay for research assistants was about $4,500 for a full year of full-time work: the majority of assistants were doctoral candidates working half-time (twenty hours per week) for nine months, for a total of $1,800. Fellowships or scholarships, for which no research work other than thesis work was required, were much fewer in number, and varied in amount. Provision of space for research appeared to be reasonably adequate for these days of heavily increased enrolments, although such space varied from modern air-conditioned premises to attic rooms of very old houses. Office space for graduate students was often insufficient. Much of educational research, utilizing ordinary classrooms and pupils, requires little additional expensive equipment; on the other hand, laboratory research often requires equipment provision paralleling the requirements of laboratory research in other disciplines. In all ten universities the facilities of digital computers were readily available to educational research workers, while less expensive IBM equipment (such as punches, verifiers, and sorters) was often on the spot for staff and student use. Well-stocked libraries were common to all the universities.

Underlying the whole matter of provision for research, of course, is the problem of finance. This rather all-pervasive problem can here only be mentioned in passing, since adequate treatment would require considerable space. It might be noted that the three main sources of finance included: (a) university sources, both regular budgets and funds to which application might be made for specific projects, (b) grants from outside agencies, for which careful proposals were required, and (c) contract research.

Summary

A rather basic conflict in the organization of educational research within universities has been noted — that of stimulating the creativity of individual professors while at the same time providing an efficient administrative structure for the specialized, complex, large, and dispersed activities which modern educational research encompasses. In the ten universities of midwestern United States visited by the writer, it was observed that a basic administrative unit consisted of the individual professor with his group of graduate students. Though bureaus of educational research served very useful purposes in some instances, in none of the institutions did it appear to be felt that educational research should be in the main turned over to such bureaus. Among the administrative structures most productive of research were large long-term organizations

whose scholarly interests were more specifically oriented, and which included research as one of their major activities in connection with those interests. Whether or not large-scale organizations existed, the temporary banding together of small groups of professors into ad hoc committees for the purpose of completing specific projects was common.

Two rather recent and decidedly promising developments were the establishment of college-wide standing committees on educational research, and the appointment of co-ordinators of educational research. Both of these arrangements were considered as extensions of the leading, co-ordinating, and facilitating functions of the Dean's office, rather than as extensions of the ultimately directing functions of that office. In most of the institutions visited certain individuals were allocated time and responsibility to serve as consultants on research design.

Certain other facilitating provisions included time for research, research assistants, space, equipment, and adequate finance. But regardless of the administrative structures and other facilities that may be planned for the purpose of encouraging efficient educational research, as Dean Cook of Minnesota observed, "Positions will build around personalities." And in the same vein Virgil Herrick of Wisconsin commented, "If a person is qualified to be on a university staff, you don't have to tell him how to, or what to do research-wise."

REFERENCE

(1) Clarke, S. C. T. "Trends and Problems in Educational Research." *Alberta Journal of Educational Research,* III, 4 (December, 1957), 209-219.

THE DEVELOPMENT OF THE ENGLISH PROGRAM IN THE SECONDARY SCHOOLS OF ALBERTA

Stanley W. Sawicki
St. Francis Boys' School, Lethbridge

Introduction

Since 1902, the English program in the secondary schools of Alberta has been revised five times. Even a casual survey would indicate that there have been great changes in the English curriculum during the last half century. The purpose of this study to describe the evolution of the English program in the secondary schools of Alberta since 1905, to point out the ways in which each program differed from those which preceded it, and to attempt to ascertain the reasons for the changes.

The English Curriculum From 1905 to 1912

When Alberta became a province in 1905, the new provincial government was content to keep in effect the curriculum which had been inherited from the Territorial government. This curriculum had been the work largely of one man — Dr. D. J. Goggin, who had been Superintendent of Education of the North-West Territories from 1892 to 1903. Just prior to his retirement, Dr. Goggin had completed a revision of the entire curriculum, wth the revisions coming into effect in 1902.

In his remarks on the introduction of the new curriculum, Dr. Goggin had defined a school as a place "where the pupil must be prepared as a member of society to live a worthy life and earn a respectable living". (9) Educators generally will agree with this definition but they disagree in their views of how this purpose can be achieved. One group claims that there is a certain body of knowledge which the pupil must learn before he takes his place as an adult member of society. According to this group, the school is, in a way, a repository of this knowledge and must discharge its responsibility of handing it on to the pupil. Those who hold the other view suggest that the child rather than the culture be taken as the starting point in determining curriculum. We should learn all we can about the child: his growth, intellectual and emotional development, his needs, and so on. Then the curriculum can be built to foster this growth.

The underlying assumption of the English curriculum during this period was that it was the school's function to transmit the cultural heritage to the students. This assumption was especially evident in the literature and reading courses. Certain works of literature had

come to be accepted as being the great masterpieces of the language and, unless a student had read and studied some of them thoroughly, he had not really been educated. Therefore, regardless of a student's abilities and interests in high school he was required to study those works which, because of their literary worth and because they had stood the test of time, had been accepted as classics. Among the prescribed selections are found such titles as: *Ivanhoe, The Merchant of Venice, Silas Marner, The Palace of Art, Locksley Hall, Il Penseroso, Lycidas, Hamlet.*

The choice of literature and the method of teaching it were determined, once the above assumption was adopted. The study of literature was thought of as a type of character training. By reading and studying the works of the "master minds", an individual would come to a realization of what was best in himself and in others and would thus develop a set of high ideals. He would read about noble actions and characters as described by the greatest authors and be inspired to emulate them.

The government examinations which were given in each standard were reprinted in the annual reports. Unlike the modern English examinations in which there is heavy emphasis on ability to read and interpret selections which are new to the student, the examinations of this early period consisted mostly of questions based directly on the material which had been studied in class. In view of this fact, there can be little doubt that teachers prepared their students for the examinations by a rigorous line-by-line analysis of the stories, poems and plays which were prescribed in the program of studies. That they were successful in this endeavor is indicated in a statement made in the sub-examiners' report in 1902: "The answer papers in poetical literature reveal a satisfactory knowledge of the subject matter and a general appreciation of its literary value. Evidently the teaching has been on sound lines." (10)

The emphasis in composition assignments was on reproduction of material which the student had studied in other subjects. Teachers were directed to give instruction in the fundamental principles of rhetoric in addition to assigning written work. In the higher standards, students were to engage in "analysis of typical passages of prose to discover their essential characteristics." (8)

Of all the subjects on the secondary school curriculum, composition received the most attention from the school inspector and the sub-examiners. These officials consistently expressed dissatisfaction with the classroom work in the subject and also with the results as revealed on the examinations. Suggestions for improving the work in composition were not lacking. While some inspectors felt that students should be encouraged to narrate their own ex-

periences rather than reproduce the substance of lessons, others advocated more formal teaching, and the sub-examiners in 1902 emphatically endorsed more practice in composition based upon the prose literature.

Grammar and rhetoric were taught in a formal manner, as a separate subject from composition. Grammar was justified on the grounds that is gave excellent training in logical reasoning and that, because of its difficulty, it built moral fibre. Dr. Goggin was of the opinion that the study of grammar made only a very limited contribution to correctness of expression. He wrote, "Correct speaking depends more on the forms of speech heard and read than on the rules of grammar, hence considerable attention is given to the correction of errors in the language used by pupils". (8)

The English Curriculum From 1912 to 1922

When the province of Alberta was created in 1905, the New Department of Education was faced with the problems of organization, and therefore found it convenient to continue with the curriculum which had been introduced in 1902. The first step towards the introduction of a new curriculum was taken several years later when a committee of "prominent educationalists of the province was appointed by the Minister of Education to consider the revision of the program of studies, and to make such recommendations as might be deemed desirable with a view to its improvement". (1)

In its first attempt at curriculum revision, the Department of Education was not prepared to go very far beyond the scope of the Goggin curriculum. Even though sources such as inspector's reports indicate that there was criticism of the curriculum and of teaching methods, the inertia of tradition was too great to be overcome in one bound.

In spite of changes in a few minor details, the 1912 program could hardly be called a new curriculum. There was a new stress on oral reading which, it was maintained, would greatly assist the student in understanding and appreciating literature. However, basic principles and emphases were unchanged; the English program was still mainly a subject-centred curriculum. In literature, for example, a knowledge of the great literature of the past was considered essential. It was the teacher's duty to see that the students read the classics, studied them in detail and learned to interpret and appreciate them. The student was required to study various types of composition and was expected to master the principles of composition and rhetoric. And when he wrote his themes, he was to base them on something he had studied in school.

The English curriculum was organized in a formal way, as the following summary of the arrangement of topics in composition

indicates. From a study of narrative composition in grade IX, the pupil would progress in each grade to other types of composition which increased in difficulty. Thus he studied description in grade X, exposition and argument in grade XI, and after reviewing these topics in grade XII, he studied the principles of rhetoric.

Why did the objective of formal intellectual discipline receive such emphasis? During the nineteenth century and the first two decades of the twentieth, the primary purpose of the high schools was to prepare students for university and normal school. It was believed that they would be better prepared for the rigors of university life if their high school studies had been difficult and disagreeable. But even if a student did not go on to a higher education, the type of training his mind had received in high school would enable him to meet and solve the problems of life more effectively.

The English Curriculum From 1922 to 1937

In the early years of the twentieth century the objectives of secondary school education in America underwent considerable change. Whereas the main function of these schools had been to prepare students for universtiy, now they had to meet the needs of the ever-increasing number of students who were attending high schools with no intention of proceeding to higher education. A growing body of literature was critical of traditional curricula and methods and was advocating changes which would better enable the school to perform its function in a changing society. Among the books published was one which was to have an important influence on English curricula—the Report of the National Committee on English published in 1917. (7)

This Report took issue with the view of English as a formal subject. It was claimed that forcing the child to pay attention to the forms of speech and writing resulted in "stultifying indifference, mere memory of words without real significance and an utter failure to establish right habits of speech or of interpretation". The feeling was growing that the detailed study of grammar and rhetoric and the doing of lengthy exercises in sentence analysis did not result in enough improvement in the students' language skills to justify the time spent on this work.

Therefore, it was advocated that the detailed teaching of the rules of grammar and rhetoric be either partially or completely abandoned. It was proposed that, since language skills must ultimately be used in social situations, they be taught in such situations. That is, the course should be made up of experiences of "social value" which would give the students practice in communication. "The pupil must speak or write to or for somebody, with a consciously conceived purpose to inform, convince, inspire or entertain."

The material for oral and written work should be taken from the experiences of the pupil—from his school work and also from his outside interests such as play, amusements, work, home activities, etc. It was recommended also that the exercises have, as far as possible, a purpose such as writing an article for the school paper. Thus the Report stressed that English must be "social in content and social in method of acquirement".

There were new trends also evident in the teaching of literature. One of these was the greater emphasis placed on interest and enjoyment. The Report (7) stated:

> Literature chosen for any given school should make a natural appeal to the pupils concerned, for without interest, which depends upon this appeal, there will be no enjoyment; without enjoyment, there will be no beneficial result.

Classic literature was still heavily favored although it was admitted that modern literature had a legitimate place in high school work. The advantage of current books and magazines was that they quickly won the interest of the students and therefore if they had a high ethical or social message, they could well be included in the curriculum.

The English curriculum which was introduced in Alberta in 1922 reflected the above ideas to some extent, but at the same time retained many traditional ideas and practices. In the literature courses, the selections to be read and studied were still drawn largely from the classics. However, for the first time since the formation of the province, a considerable number of selections from Canadian literature were placed on the program. In grade IX, selected stories from Canadian prose were prescribed, while in the grade X course, several poems by Canadian authors were to be found.

There was a new emphasis also on taking students' interests into account. "In English 1 and 2 the material is largely narrative and has been chosen because of its inherent interest for pupils of the Junior High School grades" (2). This recognition of the importance of the student's interest in his work was reflected in the composition course.

> People express themselves best about things with which they are familiar, and in which they are interested. The teacher of composition should keep this truth constantly in mind in assigning themes for pupils. For the most part, subjects should be chosen within the range of the pupil's knowledge, and as far as possible should have an individual or personal interest for him (2).

There were no separate grammar courses. Grammar was included, however, in the composition courses, ,and was to be studied in considerable detail. By integrating the study of grammer and composition in one course in each grade, it was felt that the student

would be assisted more effectively in achieving correct, effective expression.

The English Curriculum From 1937 to 1954

Until 1922, the curriculum in Alberta high schools had been rigidly prescriptive, each subject consisting of a certain well-defined body of subject matter which had to be studied and at least partially mastered by every student, regardless of his ability or interests. By 1922, however, when the curriculum was re-organized, educational thought was veering toward the principle that the starting point of any curriculum should be the student himself—his needs, interests and abilities. The findings of the ever -increasing amount of educational research which was being carried on forced educators to re-appraise traditional ideas and methods. Research into individual differences, for instance, indicated that all students should not be forced to study the same materials at the same rate. When research revealed the wide variation in the skills and abilities of students in any one class, the necessity for greater flexibility in curricula to meet the varied needs of the pupils was realized. Research in the field of grammar and language skills led to changes in attitude toward the teaching of grammar.

In 1922 a beginning had been made toward modifying the traditional curriculum, but in 1937 the break was made more definite. The needs and interests of the students were the determinant of a great part of the new program in both literature and composition. The free reading list, for instance, was greatly widened in scope so that books which had a greater appeal for young people were included. While books which had been on the curriculum for decades were still retained, new books by contemporary authors were now placed on the list. The program of studies stated, "Sufficient material has been included in each of the three courses to allow for individual differences in taste and ability" (3). In addition, the "extensive" rather than the "intensive" method of teaching literature was recommended inasmuch as it would allow for the reading and teaching of a wider variety of selections.

Another evidence of the concern which the curriculum-makers felt for meeting the varied needs of the students was the provision of a developmental reading section and the recommendation to teachers that they institute remedial programs. While even the early curricula had laid stress on student development, it was not until the period under discussion that the wide variation in rates of development of students in any one class and therefore of the need for individualizing instruction (at least to some extent) was fully recognized. The teacher was advised that the program of studies was not intended to be rigidly prescriptive and that he should

therefore adapt the course to fit the needs of the class. To assist him in this task, at least in Grade X, the textbooks assigned for literature were designed for students on two different levels of ability.

The centering of interest on the student and his needs was carried on in composition and language work. Ideally this work should grow out of his experiences, interests and problems. Topics for written and oral composition should have some meaning and significance for him. The problem of motivating the student in writing and speech activities would be lessened if he were encouraged to express himself on topics of vital concern to him.

Another significant departure from previous curricula was the combining of literature and composition into one course. The program of studies stated:

> . . . the course is set out in sections, for the sake of simplicity, but in practice these divisions should partly or wholly disappear; the English course will thus be integrated and fused into an entity so that all the activities mentioned come into play through participation in life experience.(4).

There had been a definite trend toward integration of school subjects in the United States, as evidenced by the publication of such works as *An Experience Curriculum* (6), and the practice had been adopted by many school systems. In practice, most teachers in Alberta maintained the division between language and literature. However, because the amount of literature remained considerable and the program of studies stressed reading, the ultimate effect was to reduce the amount of time spent on expressional activities.

The program also provided for a reduction in the amount of grammar to be taught and insisted that such teaching emphasize the idea of functional grammar. There is evidence that the curriculum-makers were modifying the traditional, authoritarian view of grammar toward the concept that correctness rests upon usage and that all usage is relative.

The English Curriculum As Introduced in 1954

There were six major points of difference between the courses (5) introduced in 1954 and 1955 and the ones which they superseded. Most far-reaching in its implications perhaps was the emphasis on meeting the accepted needs of youth. The language program aimed at providing the students with many opportunities to express themselves in situations and on topics which were meaningful to them and which would have a practical value for them in life. The courses provided many experiences in letter-writing, speaking, organizing and giving reports, and studying the mass modes of communication. In short, the language program was designed to pro-

vide experiences which would enable the student to become a useful, thoughtful citizen, able to express himself with competence in the many situations he would meet in life.

A second important change was the increased emphasis on the unit approach to the study of both language and literature. This method provided a practical way for the teacher to make use of the interests and experiences of the pupils as a starting point for the work in English. It also allowed greater flexibility in adapting the courses to the individual differences found amongst the members of any class and provided effective motivation for expressional activities. Planning and carrying out the work of the unit provided more opportunities for practicing democratic procedures in the classroom. In literature, the organization of the course around broad themes offered greater possibilities of attracting the interest of the pupils.

A third feature of the new program was the increase in amount of time allowed for language activities. Whereas teachers had formerly taught language for two or three periods a week, they now had five. There are several reasons why the increase in time should bring about an improvement in language skills. First is the fact that the course simply provided more time for the students to practice those skills. The teacher would also have more time to check on students' work and give more individual instruction during class periods. Because his marking burden would thus be eased, he would be free to give more assignments.

Another change in the English program was the emphasis on the communication of meaning. There was a recognition of the importance of developing motivation in language work, for students must have ideas which they want to express if they are to improve in language effectiveness. They must have something to say and should be assisted to organize and express their ideas as clearly and forcefully as possible. It was felt that in the past there had been too much emphasis on the formal aspects of grammar and too little on the significant phases of good usage.

Related to the above point was the new approach to grammar and usage. The students were to learn the grammatical concepts which would be of direct benefit to them in improving their writing and speaking. To assist them in the discussion of language problems, they were to acquire a knowledge of the terminology of grammar and usage. Grammar was to be learned in close conjunction with the compositional activities and as an aid in increasing effectiveness in those activities. To assist in this part of the work, a new textbook, *An English Handbook,*(11) was prescribed. The Handbook was not intended as a textbook on which regular

language lessons were to be based. The hope was expressed that students would develop a research technique in dealing with their individual problems in grammar and usage. Normally, the student would use the book as a reference in correcting specific difficulties which he encountered. From time to time, however, it would be necessary for the teacher to teach a lesson on some widespread difficulty.

Finally, language and literature in Grades X and XI once again became separate courses. By allowing the teacher and the class five periods a week during which they could concentrate on the task of improving language skills, the Subcommittee hoped that language efficiency would be increased. However, it was believed that the grade XII student's greater maturity would enable him to profit from an integrated course in language and literature.

Summary

1. The intensive study of relatively few literature selections by the whole class has been replaced by extensive study of a larger number of selections.
2. Research into the reading interests of young people has led to the introduction of a wider variety of types of books on high school literature courses.
3. Because educators have realized that much contemporary literature is of excellent quality as well as inherent interest to young people, many examples of this type of literature have been added to the lists of classics which have traditionally been found on English courses.
4. In language, the emphasis now is on engaging the pupils in writing and speaking activities based on their interests and experience.
5. In learning how to express themselves effectively, students must do a great deal of writing and speaking. The emphasis now is on actual practice rather than on learning the rules of composition and rhetoric.
6. Research indicates that formal study of grammar does not necessarily help students to express themselves more effectively. Therefore, the tendency today is to make grammar teaching functional, that is, to teach that grammar which will make a direct contribution to effective expression.
7. A new view of the matter of correctness in language has become current. It is generally agreed that the proper standard of correctness is the conformity of the language one uses to accepted usage rather than to its adherence to prescriptive rules of grammar.

8. There is a new stress on the developmental approach to reading. Development of reading skill is a continuous process; it cannot be assumed that reading need be taught only in the elementary grades. There are certain reading skills that the high school must teach.

REFERENCES

(1) Alberta, Department of Education. *Annual Report.* 1912, p. 81.
(2) Alberta, Department of Education. *Handbook For Secondary Schools.* 1927.
(3) Alberta, Department of Education. *Programme of Studies for the High School.* Bulletin II, 1939.
(4) Alberta, Department of Education. *Programme of Studies for the High School.* Bulletin I. 1946, p. 37.
(5) Alberta, Department of Education. *Senior High School Curriculum Guide For English.* 1955.
(6) National Council of Teachers of English. *An Experience Curriculum in English.* New York: D. Appleton-Century Company Inc., 1935.
(7) National Joint Committee On English. *Reorganization of English In Secondary Schools, Bulletin No.* 2. Washington: Government Printing Office, 1917.
(8) North-West Territories, Council of Public Instruction. *Report.* 1896, p. 22.
(9) North-West Territories, Council of Public Instruction, *Report.* 1900. p. 21.
(10) North-West Territories, Department of Education. *Annual Report.* 1902, p. 18.
(11) Scargill, M. H. *An English Handbook.* Toronto: Longmans, Green & Company, 1954.

THE ROLE OF THE PROVINCIALLY APPOINTED SUPERINTENDENT OF SCHOOLS IN LARGER UNITS OF ADMINISTRATION IN CANADA

C. P. Collins
Research Officer
The Canadian Education Association
Toronto, Ontario

SUMMARY, CONCLUSIONS, AND FURTHER RESEARCH

I—Summary

Historical Background

During the past one hundred and fifty years school systems in Canada have changed from church to local to provincially dominated systems. The most recent stage in adjustment is characterized by the development of larger units of administration.

Egerton Ryerson, while he was Superintendent of Education, developed a highly centralized system of education for Ontario. He was able to resolve the conflict between local and central control of education by delegating power to local boards to administer the *externa* and by retaining control of the *interna* under the provincial Department of Education.

The right of the provincial government to sovereign power in education was incorporated in the British North America Act of 1867. With the exception of Quebec and Newfoundland, all other provinces in Canada have adopted the system of education which was developed in Ontario. In all of these provinces the administration and supervision of the regulations concerning the interna are implemented by a corps of provincially-appointed superintendents or inspectors.

The Problem

The introduction of larger units of administration involved a significant change in the basic unit of educational administration. On the assumption that changes in administrative organization affect the obligations and behavior of the provincially-appointed superintendent of schools in larger units of administration, this inquiry is concerned with the effect which these changes had on:

1. the legal role of the superintendent as defined by statute and regulation
2. the actual role as defined by the superintendent himself
3. the amount of agreement and disagreement between the *de jure* expectations and the *de facto* behavior.

The focus of attention in the investigation of the legal and actual roles has been on the emphasis placed on the functions assigned to the position and the significance of the sources of authority.

Considerable attention has been devoted to an analysis of the organization and structure of the provincial systems of education in Canada in order to develop a conceptual frame of reference for the inquiry into the behavior of the superintendent. The position of the provincially-employed superintendent has no meaning apart from other positions to which it is related.

The Legal Role

1. *Data.* Information on the legal role of the superintendent has been obtained from school acts, annual reports of departments of education, unpublished dissertations, and histories of education.

2. *Findings.*

(1) The amount of agreement among provinces on structure, lines of authority and professional expectations is much greater than the differences.

(2) The professional expectations of the superintendent of schools, as expressed in statute and regulation, are similar in all provinces.

(3) The sources of authority expected to determine the actions and official behavior of the superintendent are the Department of Education and the board of trustees. The superintendent of schools is expected to act as a line officer for the department and as a staff officer for the board of trustees.

(4) The superintendent is expected to administer and supervise the uniform program of education for the province and to advise the board of trustees on matters which come within their jurisdiction.

The Actual Role

1. *Data.* Information on the actual behavior of the superintendent has been obtained from questionnaires distributed to all superintendents of larger units in Canada and from interviews with sixty-eight superintendents and several senior directors in departments of education in British Columbia, Alberta, and Saskatchewan.

2. *Findings.*

(1) Role Consensus.

(a) Qualifications and experience. The similarity in job specifications among the provinces is matched by similiarity in qualifications and experience reported by the superintendents.

(b) Activities. The amount of consensus is greatest in those activities which are closely related to the foundation program.

(c) Sources of Authority. Role consensus between *de jure* expectations and *de facto* perception of the significance of sources of authority is found only in a small number of tasks associated with instruction and instructional management.

(d) Functions. Role consensus between *de jure* expectations and *de facto* functional role definitions by the superintendent is found in the majority of tasks.

(2) Role Conflict. Role conflict arises from the exposure of the actor to conflicting sets of legitimate role expectations which are actual or potential sources of difficulty. Role conflict between the superintendent and the law as role definers is potential in that the superordinate authority acquiesces in the actual behavior.

(a) Sources of Authority. The superintendents of schools in each of the provinces rely more upon their own ability to influence the decisions of those with whom they work than upon authority delegated to the position.

TABLE I

PERCENTAGE DISTRIBUTION OF SOURCE OF AUTHORITY BY PROVINCE

Province	Department Directive	Board Directive	Own Initiative
British Columbia	29%	31%	40%
Alberta	20	30	50
Saskatchewan	22	24	54
Nova Scotia	30	26	44
New Brunswick	46	14	40
Canada	27	25	48

(b) Functions. The emphasis on the advisory function as a necessary part of planning indicates conditions of instability and change. The emphasis on the supervisory function and its relationship to both board and departmental policies arises from the presence of stabilized and permanent policies.

TABLE II

PERCENTAGE DISTRIBUTION OF FUNCTIONS BY PROVINCE

Province	Advisory	Supervisory	Administrative
British Columbia	47.5%	28.5%	24%
Alberta	47	33	20
Saskatchewan	50	34	16
Nova Scotia	53	26	21
New Brunswick	53	22	25
Canada	50	29	21

II—Conclusions

1. *Line and Staff Officers*

Implicit in patterns of behavior appearing in functions and sources of authority is the change in the role of the superintendent as a line officer of the Department of Education and a staff officer of the board of trustees. The superintendent advises the central authority on local needs and supervises the educational policies of the board of trustees. He advises the board of trustees and supervises the educational policies of the central authority. In his actual role the superintendent looks upon himself as a line and staff officer for both the Department of Education and the board of trustees.

2. *Balance of Power*

Superintendents of schools in Alberta and Saskatchewan are concerned about the relative amount of authority which they exercise in comparison with that of the board of trustees and the secretary of the larger unit. Superintendents of schools in Saskatchewan are particularly concerned about the lack of definition of lines of authority. The problems inherent in the balance of power relationships between superintendent and board of trustees, or their employees, are related to the conflict between *de jure* expectations and *de facto* behavior.

3. *Leadership*

The amount of consensus between the theoretical requirements of leadership and the *de facto* role of the superintendent is much greater than that between the *de jure* and *de facto* definitions of

the obligations of the superintendent as a line and staff officer. Within the frame of reference of the leadership role, the superintendent is attempting to synthesize the potential role conflicts in relationships between his position and those which are counter.

4. *Alternatives*

Resolution of the major problems depend in part, at least, on decisions to be made in relation to three alternatives:

(1) The superintendent is to remain as the line officer of the Department of Education

(2) The superintendent is to be employed by the board of trustees and become a line officer directly responsible to the local authority

(3) The position is to be legally defined in terms of leadership with formal recognition given to the right of the superintendent to act on his own initiative.

The first alternative is the present legal definition of the role of the superintendent. If this alternative is selected it should be enforced. The necessary sanctions should be employed to re-establish the office, and not the person, as most important.

If the second alternative is selected, the amount of authority exercised by the superintendent in the local hierarchy would be determined by the board of trustees. Theoretically, difficulties which are apparent in many areas would be resolved by assigning a definite status position to the superintendent. The selection of this alternative would demand a definite relaxation in central control of educational development in the province.

A decision to accept the third alternative has many implications. It would mean clear definition of the relationship between the focal position of the superintendent and such counter positions as secretary, board of trustees, principals, staff, and directors of branches within the Department of Education. It has many implications for the selection, education, in-service education and placement of superintendents. The definition of relationships between these focal and counter positions, however, needs to rest on more research findings than are available at the present time.

III—Further Research

The field of role analysis in educational administration in Canada is unexplored territory. Findings from this investigation are subject to review and reconsideration in the light of further research.

Before any definitive conclusions can be reached concerning the role of the superintendnet, consideration must be given to such problems as:

1. a role analysis of the position of the superintendent from the viewpoint of incumbents of counter positions—helping teachers, board members, secretaries of larger units, principals, teachers, directors of branches in the department of education

2. the degree to which there is consensus and conflict between the role definers in focal and counter positions

3. a role analysis of other positions as focal from the viewpoint of the superintendent.

Research is also needed to clarify the balance of power relationships among boards of trustees, Department of Education, superintendent, principal, secretary, and teachers. Examinations of expectations and behavior are needed to determine interrelationships among these positions in terms of role consensus and role conflict.

THE WANDERING QUOTIENTS SETTLE DOWN

C. C. Anderson
Faculty of Education
University of Alberta

The Problem

Dewey's 'quest for certainty' in philosophy has been paralleled in psychology by a quest for constancy and stability in scores on psychological tests. Carr (3) calls this "the constancy attitude" (3, p. 514), the tendency to assume "the existence of constants and universals" (ibid. p. 517). Both quests clearly have their origin in a desire for general and invariant laws and hence for the power of prediction which brings with it the satisfaction of control. Psychology has never been unduly blessed with a plentiful supply of such laws and psychologists have long been concerned about the apparently intractable quantitative changes or fluctuations in the function being measured, which make up part of the errors of measurement observed in psychological testing. Psychologists, of all professing scientists, have been most concerned about the extent of their errors of measurement, and perhaps only *psychological* statisticians would revert to metaphor to express their latent anxiety about it — Walker and Lev (14) state that "All measurement is infested with error" (14, p. 289) and McNemar (10) that "All measurement is befuddled with error." (10, p.294).

Indeed some authors, notably Walker (13), Loevinger (7), and Guttman (5), more concerned about the existence of fluctuation as a threat to accurate measurement and the establishment of invariant laws in psychology than with the construction of useful tests, have attempted to fashion scales which have no error variance or rather to increase as far as possible the extent to which their tests are error-free. Loevinger uses the term "homogeneous' to describe a test in which each individual passes all items up to those at a certain point of difficulty, and then fails all others of greater difficulty. His raw score will be the cumulative total of all items up to the first one that he fails. By contrast the "heterogenous" test is one in which each individual tends to sprawl his correct answers sporadically over many levels of difficulty.* Walker and Guttman have correlative terminology. However the ubiquity of errors of measurement has

*Loevinger does not indicate that the extent to which a test is homogeneous will depend partly on the fineness of grading of item-difficulty. A test with a large number of items which are only very slightly separated in difficulty would stand very little chance of being homogeneous, whereas a test of a few items at isolated points over a wide range of difficulty might be very homogeneous—and useless. Accordingly homogenity is not the sole, and not necessarily the most desirable, virtue in a test, a point which Walker recognised.

caused this type of research to peter out. Behind the rationale of the single-trial votaries lie unacceptable assumptions. For example Leovinger asserts that "Transitory variations in efficiency affect homogeneity only so far as they are tied to specific, that is, accidental contents" (7, pp. 28-29). This is based on the unwarranted assumption that errors of measurement are composed exclusively of test error. Guttman (in Stouffer et al., 11) makes the equally unwarranted assumption that quotidian fluctuation is non-existent and that therefore a single-trial coefficient is an adequate basis for the estimation of test reliability. Failure to reckon with these fallacious assumptions has been responsible for the oblivion into which 'homogeneous' tests have fallen.

However the desire to construct these tests is evidence of a vague suspicion that the existence of fluctuation implies the activity of Langmiur's (6) 'divergence', a concept which London (8) has attempted to apply in psychology.

> Convergence covers those indeterminate processes which average out into a definite state and in so doing converge to a limit and in a manner either identifiable or definable as lawful in the classic sense. Divergence, on the other hand, covers those processes which, instead of converging to a definite limit, increasingly diverge with time into a chain of events quite unrelated to expectation. (9. p.220).

Lack of convergent behaviour

> . . . prevents much of psychology and psychiatry from being exact sciences in the sense that it is always possible to set up the invariable relationship: if A, then B; and gives them more the aspect of an art than a science . . . Much of human behaviour is convergent and, hence, amenable to scientific reductions, but that which is divergent in character must remain indeterminate and be delivered over for appraisal to the psychologist's insight. (8, pp. 180-181)

It would go hard indeed with the phychometrist (and possibly even harder with the testee) if a 'true' test score were quite so indeterminate. And yet, in some recent experiments, Anderson (1) observed a steady decline in reliability coefficients with increasing time-interval (one week) between each testing. Similar evidence of apparent divergence has also been presented in another report, Anderson (2). At this point the interested spectator might very well begin to ask, as did Simpson (11) when scrutinising the spectacular increases in mental test score obtained in dubious circumstances by the Iowa environmentalist; "When are these wandering quotients going to settle down?" This plaint has considerable justification. Anderson's findings appear to conflict with the known predictive power of cognitive tests, their significant correlation with an educational criterion of success established years after the administration of the test.

A satisfactory compromise might be effected if it is hypothesised that, after an initial period of divergent decline, the reliability co-

efficients at least straighten out and begin to vary about some hypothetical mean value.

Some Preliminary Evidence

A tentative proof of this is in Anderson's (2) second report which presents for four separate cognitive function matrices of intercorrelations between parallel tests on various testing occasions. The levelling off of the reliability coefficients at approximately the sixth testing occasion can be illustrated by considering the variation in the loadings in the first (and only significant) factor. These loadings decrease in a rough progression till about the seventh testing after which they appear to oscillate. This can be illustrated by averaging each array of reliability coefficients at successive removes from the first test. Thus the array at one remove will be: r_{12}, r_{23}, r_{34}, r_{45}, r_{56}, r_{67}, r_{78}, r_{89}, $r_{9\,10}$; at two removes, r_{13}, r_{24}, r_{35}, r_{46}, r_{57}, r_{68}, r_{79}, $r_{8\,10}$, and so on. The averages of these successive arrays are listed below for the four functions involved.

TABLE I

AVERAGE RELIABILITY COEFFICIENTS AT VARIOUS REMOVES FROM THE FIRST TESTING OCCASIONS

(from Anderson, 1958b)

Experiment 1.—Verbal Intelligence									
Removes ...	1	2	3	4	5	6	7		
Ave. r	.889	.902	.887	.869	.864	.834	.860		
Experiment 2.—Verbal Intelligence									
Removes.....	1	2	3	4	5	6	7	8	9
Ave. r	.699	.663	.667	.646	.650	.513	.652	.677	.675
Experiment 2.—English									
Removes.....	1	2	3	4	5	6	7	8	9
Ave. r	.797	.820	.766	.776	.766	.702	.643	.634	.768
Experiment 2.—Arithmetic									
Removes.....	1	2	3	4	5	6	7	8	9
Ave. r	.795	.808	.784	.759	.730	.714	.745	.739	.769

These figures have illustrative value only; they are not exact because increasingly fewer correlations are being successively averaged, making the successive means less stable.

Exactly why this levelling off should occur at this particular time and place may be the subject of guesswork. Certainly, it seems clear that repeated psychological measurements show both divergent and convergent characteristics and the relative attitude to test results is the only possible one. As Dingle (4) has said:

> . . . if we define our operations so that they include only a finite number of repetitions, we must expect that the relations we find between their results will not be final ones . . . We are then faced with prescribing the operations so that they yield results whose relations are as comprehensive and lasting as possible, and it is an obvious addition to our specification to say that the number of repetitions must be at least sufficient to leave us in no doubt about the simplest smooth curve to be drawn. (4, p.330)

Conclusion

From the present evidence it might be concluded that the testing should go on for at least seven occasions, but this is not a definitive solution. There are tentative signs in the matrices that the curve of the size of reliability coefficients may be on the upswing and it is not at all impossible that as few as four testings could bring about for each subject as stable a quotient as might be extracted from the human situation. For the moment there can be no settled opinion about this until a larger number of repeated testings is made and there is 'no doubt about the simplest smooth curve to be drawn'.

REFERENCES

(1) Anderson, C. C. "Function fluctuation." *British Journal of Psychology,* Monograph Supplement, No. XXX, 1958.

(2) Anderson, C. C. "Further evidence on cognitive fluctuation." *Alberta Journal of Educational Research,* 4, 1958, 40-49.

(3) Carr, E. H. "The quest for constants." *Psychological Review,* 40, 1933, 514-532.

(4) Dingle, H. *The Scientific Adventure.* London. Pitman, 1950.

(5) Guttman, L. "A basis for analysing test—retest reliability." *Psychometrikar,* 11, 1945, 81-95.

(6) Langmuir, I. "Science, common sense and decency." *Science,* 97, 1943, 1-7.

(7) Loevinger, J. E. "A systematic approach to the construction and evaluation of tests of ability." *Psychological Monographs,* 61, 1947, No. 4.

(8) London, I. D. "Some consequences for history and psychology of Langmuir's concept of convergence and divergence of phenomena." *Psychological Review,* 52, 1946, 170-188.

(9) London, I. D. "The role of the neutralised symbol in psychiatry." *Journal of General Psychology,* 40, 1949, 229-245.

(10) McNemar, Q. "Opinion-attitude methodology." *Psychological Bulletin,* 43, 1946, 298- 374.

(11) Simpson, B. R. "The wandering I.Q.. Is it time for it to settle down?" *Journal of Psychology,* 7, 1939, 351-367.

(12) Stouffer, G. A. et al. *Measurement and Prediction,* Princeton, New Jersey: Princeton University Press, 1950.

(13) Walker, D. A. "Answer pattern and score scatter in tests and examinations." *British Journal of Psychology,* 30, 1940, 248-260.

(14) Walker, H. M. and Lev. J. *Statistical Inference.* New York: Holt, 1953.

THE EFFECT OF GROUPING ON VARIABILITY IN ACHIEVEMENT AT THE GRADE III LEVEL

S. C. T. Clarke
Faculty of Education
University of Alberta

The Problem

Learning theory supports the view that children learn most effectively when taught at a level appropriate to their maturation and readiness. Studies of individual differences show wide variability in achievement at any grade level, increasing from lower to high grades. Ideally, these two conditions would lead to individual instruction, but economy dictates that children be taught in groups. These facts pose a problem which school administrators have attempted to solve by acceleration and retardation, enrichment, and various forms of grouping.

This paper deals with reduction in the variability of achievement produced by grouping on the basis of intelligence and on the basis of reading achievement. The practical administrative problem was twofold: (1) what was the variability of achievement in existing classrooms? (2) to what extent would it be reduced by either of the two grouping procedures?

Related studies by Wrightstone (5), Cook (2), Dunlop, Hunka and Zingle (4) are reviewed by Dunlop (3) in an article which presents further evidence on the variability of achievement and ability in Grade VII and VIII pupils in groups of classroom size. He found that the range of I.Q. scores was 45 to 50 points, and the spread of grade scores in achievement was typically six grades. In groups of classroom size, selected on the basis of intelligence, variability was reduced in reading, spelling, and language by 43, 50,, and 66 per cent respectively. When reading scores were used for grouping, reduction in range of grade scores was respectively 54, 67.8, and 50.7 per cent.

The present study repeats the above investigation at the Grade III level, using actual classroom groups in five large schools.

Procedure

In the last two weeks of May, 1956, all of the Grade III pupils in the Edmonton Public Schools were given a battery of tests as part of a large-scale study. The resulting data, for five large schools including 21 Grade III classrooms (four schools each had four classrooms and the fifth five) with 621 pupils, were analyzed in the present study. The grade-score norms of the achievement tests,

based on United States populations, were used as the unit of achievement.

Variability in ability and achievement was measured in several ways. For each classroom, the range of scores was found, the middle 80 per cent of these scores, the standard deviation, and the frequency distribution by half-year intervals. With respect to any one variable such as total achievement, there were 21 ranges, 21 "middle 80 per cents", and 21 frequency distributions. To determine a typical or representative classroom, each of these measures of variability was averaged. Standard deviations were also averaged. First, a test was made for California Achievement Test Battery Total Score in the following manner: variance for each classroom was calculated, summed, averaged by dividing by the total number of students, and the square root taken. This figure was 0.43 grade scores, that is, 4.3 months of school progress for an average child. This figure was then compared with the average of the 21 standard deviations, which was 0.41 grade scores. The standard error of the difference was calculated and found to be 0.02 grade scores. From this point on, the average standard deviation was used without repeating this test for each variable.

The following points, arising out of the above procedure should be noted.

1. The average grade score at the end of May of a school year should have been 3.9 (9 months in the 3rd grade).

In fact, because United States norms as provided by the tests were used, the average grade scores were much higher. This inflation of scores should not affect interpretations of relative gains.

2. In grouping classes, crude test error, made up as pointed out by Anderson (1), of test unreliability and function fluctuation, will cause some pupils to be misplaced. Whether the true variability in classes would be greater or less than that found in this study, because of the statistical regression phenomenon associated with grouping when test error is present, is not known. Further study is indicated here.

3. This study is based on large schools, with a minimum of four classrooms per grade. While the same general results might be expected of schools with two classrooms per grade, there is no certainty that these results would follow. Again, these results are calculated at the end of Grade III, and may not apply equally well at other grade levels.

Variability in Achievement in Ungrouped Classrooms

Each of the five Principals was asked how the students in his Grade III classes had been assigned to classrooms. All Principals indicated that they had originally been assigned to classes in Grade

I to avoid homogeneous grouping. It was plain from the Principals' replies that, as students dropped out or were added, an effort had been made to avoid ability grouping and to retain equal classes, as illustrated in the following quotation:

> Criteria for grouping were: an even distribution of the average pupils; no concentration of bright or weak pupils in any one classroom; a balance of boys and girls; noisy or talkative pupils were separated; and failing pupils were usually placed with a different teacher.*

The variability of ability and achievement found in this study should therefore be typical of ungrouped classrooms.

The variability in intelligence, as measured by the California Short Form Primary Maturity Tests is shown in Table I.

TABLE I

VARIABILITY IN INTELLIGENCE IN AN AVERAGE GRADE III CLASSROOM

Average Range in I.Q.			Average and Rounded Frequency by I.Q. Interval									S.D. in I.Q.
Low	Mid 80%	High	50-59	60-69	70-79	80-89	90-99	100-109	110-119	120-129	130-139	140-149	
77	89 116	127	0.1	0.14	0.57	3.1	5.9	9.3	7.0	2.1	0.48	0.1	12.2
Rounded Frequency			....	....	1	3	6	9	7	2	....	....	

The lowest I.Q. in each of the 21 classrooms was determined, and the average of these figures was 77. Similarly, the average of the 21 highest I.Q.'s was 127. Thus, in a typical or average classroom, the range of I.Q. found was exactly 50 I.Q. points. From Table I, it can also be noted that 80 per cent of the I.Q.'s of pupils in a typical Grade III classroom will lie between 89 and 116 inclusive, and that the standard deviation of the I.Q. is 12.2. The actual average distribution of pupil scores of the 21 classrooms from Table I shows that 0.1 pupils have an I.Q. in the range 50-59, 0.14 pupils have an I.Q. in the range of 60-69, 0.57 pupils have an I.Q. in the range of 70-79, etc. Rounding these figures, we find that in typical Grade III classrooms there would be one pupil with I.Q. 70-79, three pupils with I.Q. 80-89, six with I.Q. 90-99, nine with I.Q. 100-109, etc., in a classroom with 28 pupils.

*Personal communication from David Cooney, Principal, Hazeldean School.

The variability in achievement in ungrouped classrooms is shown in Table II. As before, the lowest grade score in Word Recognition was determined for each of the 21 classrooms and averaged. Thus in the typical or average classroom the low grade is 2.7 (seven months in Grade II) and the high is 6.2. The middle 80 per cent

TABLE II

VARIABILITY IN ACHIEVEMENT IN AN AVERAGE GRADE III CLASSROOM

Test	Grade Score Units				
	Average Range				S.D.
	Low	Mid 80 %		High	
Gates Advanced Primary Word Recognition	2.7	3.2	5.4	6.2	0.93
Gates Advanced Primary Paragraph Meaning	2.8	3.3	6.0	6.9	1.06
California Achievement Tests (Primary) Reading	3.5	4.0	5.3	5.7	0.57
California Achievement Tests (Primary) Arithmetic	3.8	4.3	5.2	5.5	0.42
California Achievement Tests (Primary) Language	3.3	3.8	5.0	5.4	0.51
California Achievement (Primary) Total Test Battery	3.7	4.2	5.1	5.4	0.41

TABLE III

DISTRIBUTION OF ACHIEVEMENT IN AN AVERAGE GRADE III CLASSROOM

Test	Grade Score Interval										
	1.5—1.9	2.0—2.4	2.5—2.9	3.0—3.4	3.5—3.9	4.0—4.4	4.5—4.9	5.0—5.4	5.5—5.9	6.0—6.4	6.5—6.9	7.0—7.4
Gates Word Recognition	....	0.52 1	1.9 2	4.7 5	6.4 6	5.4 5	4.6 5	2.4 2	2.6 3	0.95 1	0.38	0.04 .
Gates Paragraph Meaning	....	0.14	1.8 2	3.6 4	6.3 6	8.5 9	2.3 2	2.1 2	*	3.3 3	0.95 1	0.57 1
California Reading	....	0.05	0.14	0.57 1	2.24 2	6.8 7	9.8 10	7.1 7	2.8 3	0.05	....	....
California Arithmetic	....	....	0.01	0.14	1.2 1	4.7 5	11.1 11	9.8 10	0.71 1	....	....	....
California Language	0.05	0.01	0.01	1.2 1	3.0 3	9.7 10	10.2 10	4.6 5	0.52 1	....	....	....
California Total Battery	....	....	0.09	0.48	1.2 1	5.9 6	12.2 12	8.0 8	0.9 1	0.14	....	....

*Zero frequencies are an artifact of the test. A raw score of 26 yields a grade score of 5.4 and a raw score of 27 a grade score of 6.0.

of the pupils score between grade 3.2 and grade 5.4,, these figures being determined as before by averaging the results from 21 classrooms. The average standard deviation is 0.93 grade scores in Word Recognition. The remainder of Table II is read in a similar manner.

Variability in achievement can be illustrated more directly in the average and rounded distributions shown in Table III. The distribution for each classroom was determined, and the average distribution for each class interval was obtained by adding the frequency for each classroom and dividing by 21. Thus, in the California Total Battery (composite scores from the Reading, Arithmetic, and Language tests), in the grade score interval 2.5 to 2.9 there were 0.09 pupils (on the average), while in the next interval there were 0.48 pupils, and in the interval grade 3.5 to grade 3.9 there were 1.2 pupils, on the average. In the space below, these fractions of pupils have been rounded off, so that from Table III one may expect the following distribution of pupils for given grade levels of total achievement in reading, arithmetic and language: grade 3.5 to grade 3.9, one pupil; grade 4.0 to 4.4, six pupils; grade 4.5 to 4.9, twelve pupils; grade 5.0 to 5.4, eight pupils; and 5.5 and above, one pupil. The remainder of Table III is read in the same manner.

The following conclusions are made about the variability of ability and achievement in ungrouped classrooms at the end of Grade III:

1. The range of intelligence is 50 I.Q. points, and approximately two-thirds of the students in a class will be in a range of 24 I.Q. points centering on the average (mean $\pm$ S.D.).

2. Ranges of Word Recognition and Paragraph Meaning are 3.5 and 4.1 grades respectively, and approximately two-thirds of any class will be within one grade of the class average (S.D's 0.93 and 1.06).

3. Range of California Reading scores is 2.2 grades, and approximately two-thirds of any class will be within 6 months of the class average (S.D. 0.57)

4. Variability in arithmetic is lower than in reading or language.

5. Range of California Language scores is 2.1 grades, and approximately two-thirds of any class will be within half a school year (5 months) of the class average (S.D. 0.51).

The Effect of Grouping

In each of the five schools, the classrooms were re-grouped on the basis of intelligence. The pupils were ranked by I.Q., and split into classrooms of equal size. For these new classrooms, now grouped by intelligence, the previous calculations were repeated.

The same procedure was followed with respect to achievement in

reading as measured by the California Reading Test. The classrooms were regrouped on this basis and variability again calculated.

Table IV shows the effect on variability in intelligence of grouping by California Reading achievement scores. Range is reduced from 50 I.Q. points to 45 I.Q. points, or 10 per cent; the standard deviation is reduced from 12.2 to 11.1 I.Q. points or 9 per cent. Perhaps these data are of only academic interest, since it is variability in achievement which is of more concern to principals.

TABLE IV

EFFECT OF GROUPING BY READING ON VARIABILITY IN I.Q.

	Av. Range in I.Q. Scores				Average Frequency in I.Q. Interval									S.D. in I.Q. Scores
	Low	Mid	80%	High	50-59	60-69	70-79	80-89	90-99	100-109	110-119	120-129	130-139	140-149	
Uugrouped	77	89	116	127	0.1	0.14	0.57	3.1	5.9	9.3	7.0	2.1	0.48	0.1	12.2
Grouped by Reading	82	91	115	127	0.06	0.06	0.59	2.9	5.9	10.1	6.5	1.6	0.59	0.06	11.1

TABLE V

EFFECT OF GROUPING ON WORD RECOGNITION GRADE SCORES

	Av. Range in Grade Scores				Average Frequency by Grade Score Interval										S.D. in Grade Scores
	Low	Mid	80%	High	2-2.4	2.5-2.9	3.0-3.4	3.5-3.9	4.0-4.4	4.5-4.9	5.0-5.4	5.5-5.9	6.0-6.4	6.5-6.9	7.0-7.4	
Ungrouped	2.7	3.2	5.4	6.2	0.52	1.8	4.7	6.4	5.4	4.6	2.4	2.6	0.95	0.38	0.04	0.93
Grouped by I.Q.	2.7	3.2	5.3	6.1	0.38	1.6	4.5	6.0	5.5	4.2	2.1	2.6	0.95	0.33	0.05	0.88
Grouped by Reading	2.9	3.4	4.5	5.4	0.50	1.8	5.3	7.5	4.3	3.0	0.75	2.0	0.25	0.25	....	0.58

Table V indicates the effect of grouping by I.Q. and by reading on Gates Advanced Primary Word Recognition grade scores. Range

of scores for ungrouped classrooms was 3.5 grades. When grouped by I.Q. this range dropped to 3.4 grades (3 per cent), and as could perhaps be expected, when grouped by reading the range dropped to 2.5, or by 29 per cent. Standard deviation of grade scores was 0.93 for ungrouped classrooms and dropped to 0.88, or by 5 per cent, when the classes were grouped by I.Q. The standard deviation dropped to 0.58, or by 38 per cent, when classes were grouped on the basis of reading achievement.

TABLE VI
EFFECT OF GROUPING ON CALIFORNIA READING GRADE SCORES

	Av. Range in Grade Scores				Average Frequency by Grade Score Interval								S.D. in Grade Scores
	Low	Mid	80%	High	2.0-2.4	2.5-2.9	3.0-3.4	3.5-3.9	4.0-4.4	4.5-4.9	5.0-5.4	5.5-5.9	6.0-6.4	
Ungrouped	3.5	4.0	5.3	5.7	0.05	0.14	0.57	2.2	6.8	9.8	7.1	2.8	0.05	0.57
Grouped by I.Q.	3.4	4.0	5.3	5.6	0.05	0.09	0.71	1.7	6.6	8.9	7.0	2.9	0.05	0.52

TABLE VII
EFFECT OF GROUPING ON CALIFORNIA ARITHMETIC GRADE SCORES

	Av. Range in Grade Scores				Average Frequency by Grade Score Interval								S.D. in Grade Scores
	Low	Mid	80%	High	2.5-2.9	3.0-3.4	3.5-3.9	4.0-4.4	4.5-4.9	5.0-5.4	5.5-5.9	6.0-6.4	
Ungrouped	3.8	4.3	5.2	5.5	0.01	0.14	1.2	4.7	11.1	9.8	0.71	....	0.42
Grouped by I.Q.	3.9	4.4	5.2	5.5	0.09	0.09	0.67	3.6	11.0	11.7	1.1	0.05	0.41
Grouped by Reading	4.0	4.4	5.2	5.4	0.06	0.18	0.76	4.1	11.4	11.7	0.72	....	0.33

Table VI presents the effect of grouping by intelligence on variability in California Reading. The range of reading scores is not reduced, but remains at 2.2 grades. The standard deviation is re-

duced from 0.57 to 0.52 or by 9 per cent. An examination of the distribution also shows the small effect on frequencies of pupils obtaining given grade scores produced by grouping on the basis of I.Q.

Table VII shows the effect of grouping by intelligence and by reading on California Arithmetic grade scores. The range of arithmetic grade scores for ungrouped classrooms was 1.7 grades and was reduced by grouping by I.Q. to 1.6 grades, or by 6 per cent. The standard deviation of grade scores was reduced by grouping on the basis of I.Q. by 2 per cent. Grouping by reading reduced the range from 1.7 grades to 1.4 grades or 18 per cent, and reduced the standard deviation from 0.42 to 0.33 or 21 per cent. This rather interesting finding indicates that grouping by reading is markedly superior in reducing variability in achievement in arithmetic at this grade level.

TABLE VIII

EFFECT OF GROUPING ON CALIFORNIA LANGUAGE GRADE SCORES

	Av. Range in Grade Scores				Average Frequency by Grade Score Interval									S.D. in Grade Scores
	Low	Mid	80%	High	1.5-1.9	2.0-2.4	2.5-2.9	3.0-3.4	3.5-3.9	4.0-4.4	4.5-4.9	5.0-5.4	5.5-5.9	
Ungrouped	3.3	3.8	5.0	5.4	0.05	0.01	0.01	1.2	3.0	9.7	10.2	4.6	0.52	0.51
Grouped by I.Q.	3.2	3.8	5.0	5.4	0.05	0.01	0.01	1.1	2.8	8.9	9.9	4.3	0.62	0.50
Grouped by Reading	3.5	4.0	4.9	5.2	0.06	0.12	0.12	1.1	2.9	9.5	10.2	4.7	0.67	0.41

Table VIII presents the effect of grouping by intelligence and by reading on California Language grade scores. Calculated as before, the reduction in range produced by grouping by I.Q. is one month in grade scores, or 3 per cent; while the corresponding reduction in standard deviation is 2 per cent. Grouping by achievement on the California Reading test reduces the range in California Language scores from 2.1 grades to 1.7 grades, or 20 per cent, and reduces standard deviation from 0.51 grade scores to 0.41, or 20 per cent.

The following conclusions are made about the effect of grouping on variability in ability and achievement in classroom groups at the end of Grade III.

1. Grouping by California Reading test scores reduces variability in California Mental Maturity I.Q. by 10 per cent.

2. Grouping by I.Q. reduces variability in California Reading test scores by 9 per cent.

3. Grouping by I. Q. reduces variability in arithmetic and language by 2 per cent, while grouping by reading reduces the variability by 20 per cent.

4. Obviously, grouping by reading reduces variability in word recognition and in paragraph meaning much more than does grouping by I.Q.

Discussion

The practical problem facing school administrators of large schools is whether to group or not to group. The arguments pro and con have been stated many times. If the administrator decides to use grouping by some sort of measure, then the problem becomes: which measure? The results of the present study would favor a reading test over an intelligence test, if the purpose of the grouping is to reduce variability in achievement.

This study finds that the extent of reduction in variability which can be achieved by grouping is not great. The most encouraging finding is that grouping by reading, which would obviously reduce variability in reading by a maximum amount, also reduced variability in both arithmetic and language achievement scores by twenty per cent.

This study corroborates the findings of previous research, in that individual differences in achievement in any classroom are great, but not so great at the Grade III level as in higher grades. The reduction in variability in achievement resulting from grouping by intelligence was lower than usually found. Perhaps the non-language portion of the California Mental Maturity tests used is less related to success in academic work than is the case with items used in other intelligence tests. The reduction in variability in achievement resulting from grouping by reading scores was the same as that found by Cook (2). However, this reduction in variability is considerably less than that reported by Dunlop (3). The differences in results may be caused by the lower grade level of the present study, and by the use in this study of average standard deviation as a measure of variability.

The writer would propose the following further studies which could be the basis for, or result from, administrative try-out.

1. Use a different intelligence test for grouping to determine whether greater reduction in variaability of achievement would be found.

2. Try a combination of reading scores and I.Q. scores as a basis of grouping, to determine whether the combination of these measures would reduce variability more than does either one alone.

3. Test the effect of statistical regression by re-testing classes grouped by any of the above methods.

REFERENCES

(1) Anderson, C. C. "Function Fluctuation." *British Journal of Psychology,* Monograph Supplement No. 30, 1958.

(2) Cook, W. W. "Grouping and Promotion in the Elementary School." *Series on the Individualization of Instruction,* Number 2, University of Minnesota, 1941.

(3) Dunlop, G. M. "Further Evidence on the Control of Individual Differences in the Classroom." *Alberta Journal of Educational Research,* June, 1957.

(4) Dunlop, G. M., Hunka, S. and Zingle, H. "Individual Differences in Alberta Schools." *Alberta Journal of Educational Research,* December, 1955, 5-15.

(5) Wrightstone, J. W. "Analysis of Intelligence and General Achievement, Third Grade Classes, New York City." March-May 1949, Final Summary Report, *Bureau of Reference, Research and Statistics,* New York City Schools.

A NOTE ON THE USE IN ALBERTA OF THE SEQUENTIAL TESTS OF EDUCATIONAL PROGRESS: ESSAY TEST

DONALD B. BLACK
Faculty of Education

The STEP Essay Test

The past year has seen the release of the Cooperative Test Division's STEP Essay Test, the return of a valuable evaluating device to its rightful place in the school's testing program.

Critics have not been kind to the essay or free response test. With the advent of the objective-type test, it was shown rather conclusively that the essay test had low validity, low reliability, and was time-consuming to score properly. There was little that could be shown for the essay test that could not be shown for the then new objective tests.

Not all educators have been convinced, and many have remained stout advocates of the usefulness of the essay. They point out that one has only to go into any classroom to see that the essay is undoubtedly the most popular type of test used by the teacher. Stalnaker (8) points out, however, that the essay test has remained essentially undeveloped and lacking in needed research and, therefor, many of its qualities have never been assessed.

One organization has continued research on the essay test since the late twenties. This is the CEEB (the College Entrance Examination Board) (3), and the reader is referred to almost any back issue of two publications released by this organization, the College Board Review and the Annual Report of the Director, for discussions of essay test research undertaken by this organization. In 1948, the test development activities of the CEEB were merged with those of several other similar non-profit organizations to form Educational Testing Service (6) where much of the essay test research was continued. The Cooperative Test Division, publishers of the STEP Essay Tests, is part of this organization.

Diederich (5) summarizes current research in essay examinations conducted at Educational Testing Services by pointing out that the most recent findings seem to establish that essay tests cannot be made either sufficiently reliable or sufficiently valid to be used as a basis for admission to college. He is careful to use the word "sufficiently" in his remarks to point out a degree of adequacy for a specific purpose.

In his paper, Diederich summarizes recent essay research findings as follows: (1) Verbal intelligence (objective) is a better pre-

dictor of a teacher's rating of a student's written work for a year on *mechanics, style, organization, content,* and *reasoning* than the equivalent score on an essay test, specifically, the CEEB General Composition Test. (2) The correlation between ratings of students who wrote the same essay test twice is about .45. (3) A verbal intelligence test score picked out the good and poor writers better than the essay test and with better consistency. (4) A one hour objective English test score agreed more closely with the teacher's rating than did the essay test. (5) The essay score depended more upon the good-luck choice of a particular essay topic when several alternatives were available than upon any other factor that could be found.

Diederich adds that even under the most ideal of circumstances, one cannot hope for a group of readers to have much higher than a .70 inter-reader reliability. To achieve even this, Diederich proposes five principles for scoring essay tests:

(1) Every piece of identifying data about the student should be removed from the answer sheet.

(2) Any paper that will make a serious difference in a student's grade should be read twice, and if these two grades do not agree they should be read a third time by a third reader.

(3) It is desirable to get at least two papers from each student, preferably in different testing situations.

(4) Grading standards of the different readers can be made more nearly comparable by having the readers sort papers into a given number of piles in order of merit, and to put a certain proportion of papers in each pile.

(5) Evaluate on how well the essay examinations are working. Not only talk out differences between staff, but be alert for readers who consistently over- or underrate.

Diederich, after working on this problem for many years, sums up his position as follows:

> Even the best essay examinations and even the most expert readers using the best methods thus far devised, can achieve only dubious validity and reliability, and I hate to think what either the validity and reliability of the average essay examination must be, if it has never been submitted to critical analysis . . . Until you have subjected your own examinations to some of the simple checks . . . found out how sharply others disagree with your judgments, and even how you disagree with yourself on different occasions, you will never find out how fallible your grades are, and hence will have no incentive to improve and no *method* for measuring your improvement.

The writer has quoted freely from Diederich's paper because it ably reflects the remarks of others in the field. For example, from five test-and-measurement texts, chosen at random (Ross and Stanley (7), Thorndike and Hagen (9), Torgerson and Adams (10), Bradfield and Moredock (2), and Baron and Bernard (1), it was found

that all cover several or all of the points made by Diederich, particularly in regard to the improvement of grading practices. However, Diederich's paper was chosen for still another reason. The Cooperative Test Division's STEP Essay Tests incorporate many of the features recommended by Diederich as the result of the essay research programs in the Research Division of the same organization.

The Essay Tests were published in 1957 as part of the Cooperative Test Division's new test battery, Sequential Tests of Education Progress (STEP). In all, four forms of the test at each of the four levels are available. The grade ranges of the four levels are 1: Freshman and Sophomore College (Grades 13 and 14); Level 2: Grades 10-12; Level 3: Grades 7-9; Level 4: Grades 4-6. Each test is thirty-five minutes long.

These tests have the following features:

(a) Students write on single assigned topics.

(b) Raters divide their rated essays into seven groups or piles.

(c) Comparison essays are provided as "bench-marks" for groups 2, 4, and 6 of the seven groups to which the reader refers the paper being read as being equal to, better or poorer than, before classifying.

(d) Readers are instructed to score papers on the basis of quality of thought, style, and conventions. They are cautioned to avoid "marking down" or penalizing a student for illegibility in writing, general untidiness, "disputable" grammatical errors, and errors in spelling. These are considered extraneous to the value of the essay as a communication of ideas. Though these may be important, their evaluation is not provided for in evaluatory performance on the Essay Test.

(e) Readers are cautioned against the "halo effect" in rating, and are urged to adopt procedures that will ensure anonymity of the writer.

(f) The instructions suggest two readings by different readers for each paper, with neither reader knowing what mark the other gave as a score. When ratings diverge greatly, another reading by a third party is recommended to resolve the issue.

(g) The seven-point rating scale is convertible to scale-score units consistent with the other tests of the STEP series. Each form of each level of the Essay Test has a different conversion scale.

The publisher in his Handbook for the Essay Tests (4:3) points out the following nine purposes of the tests:

1. Assisting teachers to check their judgments of the ability to write through the use of common essay topics and comparison

papers. It is noted that the comparison essays are samples not of what experts might expect or hope for as standards, but of what does exist in many schools.

2. Identifying students who are especially advanced or retarded in free-writing ability, and studying students' development in writing ability.
3. Furnishing a basis for comparing free-writing performance of classes, schools, and other groups.
4. Providing a continuous scale of free-writing ability, covering grades 4 through 14.
5. Furnishing teachers with standardized topics for free-writing exercises.
6. Promoting essay writing by the teaching-towards-testing idea; i.e., influencing instruction because of anticipation of tests of this sort.
7. Providing, by example, help for teachers in devising topics which are interesting and challenging to students and, as a consequence, promote better writing.
8. Influencing teaching towards the important characteristics of good writing; e.g., stressing the relative importance of quality of thought, style, and conventions.
9. Serving as a basis for class discussion of the criteria of good writing.

For detailed information concerning the development of these tests (4:4), the criteria for choosing essay topics (4:4), and detailed norming and standardized data (4:26-32), the reader is referred to any one of the four Handbooks for Essay Tests. The topics were selected by a committee of nationally recognized American experts in curriculum and methods of language teaching. The tests were administered to 5,000 students in sixty institutions. Each student wrote at least two of the Essay Tests and all wrote the appropriate level of the School and College Ability Test (SCAT). From this sample, the comparison essays were selected and the papers reread twice by different raters. The norms, however, are not claimed to be based upon a representative sample of the entire United States (4:26).

The Clover Bar Testing Program

From A of all levels of the STEP Essay Test was administered in March, 1958, as part of the testing program of an educational survey being conducted in the Clover Bar School Division. This school division borders Edmonton and extends into the farmland east and south-east of the city. Just under fifteen hundred students in grades 4 to 12 wrote one Essay test each.

Tests were not scored by the classroom teachers although they were administered by them. The first reading was made by the senior students in two University of Alberta Education classes, (Tests and Measurement and English Methods). The second readings, and where necessary, third readings, were done by a small team of seniors, graduate students and faculty members.

Marking of the papers was rigorously supervised. Each reader was provided with an Essay Test Handbook at the level appropriate to the Essay Tests he was scoring. A briefing session was held in every instance before scoring and strict adherence to the test scoring directions in the Handbook (4:10-20) emphasized.

For the statistical analysis of these papers, only the ratings of the first two readers were used. Further, all marks reported in this paper are in rating units (scale of seven) rather than in scale score units of STEP.

In all cases, the tests were pooled by level for scoring regardless of the grade indicated. Although each student was identfiied by number, no attempt was made to conceal the identity of the writer of each paper as no reader knew any member of the Clover Bar staff or any of the pupils of the Clover Bar School Division. Marks on the first reading were concealed from the second reader, and were revealed only after the second reading. At this time, the papers were examined to discover those with grading variances of three grades or more. These papers were then read by a third reader. During this third reading, all previous marks were again concealed.

This paper, then, seeks to consider the reader-reliability of the scoring of these tests and to assess the variability of the score returns in comparison with the data provided by the publisher. Since there is always some doubt in the minds of Canadian educators as to the suitability of American tests and test data for Canadian schools, this paper may, in part, answer this doubt. It must be noted, however, that the evidence is based upon the use of these tests in only one Alberta school division.

The first statistics of interest examined were the inter-reader reliability coefficient of the reading of the Essay tests, i.s., how consistent were the local readers in their rating of the same paper? The product moment correlation coefficients between the first two readings were calculated. The publisher provides similar data in the Handbook. The findings of this analysis are reported in Table I.

This Table shows that except for Grades 5 and 6 on the level 4A Essay Test, the reliability coefficients for the local reading of the papers were higher than those reported by the publisher for the Level concerned. When the Clover Bar data are combined for comparison by test level only, the reliability coefficients for local

scoring are higher than those reported by the publisher by level, although it is obvious that no real difference exists between reported reliability coefficients for Level 4.

TABLE I

COMPARISON OF READING RELIABILITY OF STEP ESSAY TEST BY GRADE FOR CLOVER BAR STUDENTS

Essay Test		Clover Bar Grade	Clover Bar Students		Publisher's Data	
Level	Form		r.	N.	r.	N.
1	A	12	.60	68	.50	197
2	A	10	.79	147		
		11	.84	99		
		combined	.82	246	.68	376
3	A	7	.80	224		
		8	.82	229		
		9	.80	199		
		combined	.81	652	.74	399
4	A	4	.76	295		
		5	.66	274		
		6.	.68	245		
		combined	.71	814	.70	256

Further analysis dealt with the variance of the reader scores. The variability of the Clover Bar marks was determined in two ways. First, by using the sum of the scores of the first two readings to determine the standard deviation, and from this an estimated standard deviation for the test was calculated by reducing the test to half length. Second, the average of the two scores rounded to the nearest whole unit was used to determine the standard deviation for the reported scores. The standard errors of measures for these two methods of measuring variability of marks were also calculated. However, apart from the inter-reader reliability coefficients, it was not possible to calculate the score reliability coefficients since Clover Bar students wrote only one Essay Test. It was necessary, then, to use those reported by the publisher. It should be noted that these score reliability coefficients are always lower than inter-

reader reliabilties because they are influenced by the variations not only in reader's grades, but in student performance as well. The findings from this analysis are summarized in Table II.

TABLE II

RELIABILITY AND SCORE VARIANCE STATISTICS FOR STEP ESSAY TESTS ADMINISTERED TO CLOVER BAR STUDENTS

Essay Test		Grade	r.	S_1	S_2	S.E. $(meas.)_1$	S.E. $(meas.)_2$	S.E. $(meas.)_p$
Form	Level							
1	A	12	.50	2.58	.96	1.59	.68	.88
2	A	10		2.89	1.07	1.77	.66	
		11		2.67	1,26	1.64	.78	
		combined	.62	2.90	1.16	1.78	.72	.87
3	A	7		2.60	1.09	1.60	.67	
		8		2.63	1.05	1.68	.65	
		9		2.73	1.05	1.60	.65	
		combined	.62	2.76	1.05	1.70	.65	.79
4	A	4		2.67	.97	1.81	.66	
		5		2.67	1.02	1.81	.69	
		6		2.71	.89	1.84	.60	
		combined	.54	2.74	1.03	1.85	.70	.87

Where:

r = estimated reliability of the Essay Test by Level for single reading as reported by publisher (4:30).

$S_1 = \frac{S_c + \sqrt{2}}{\sqrt{1-r}}$ is standard deviation of one reading where S_c is the standard deviation for the sum of two readings and r is the reliability of a single reading.

S_2 = is the standard deviation of the distribution of the resolved or assigned scores from two readings of the same papers.

S.E. $(meas.) = S\sqrt{1-n}$ where S is the standard deviation of the distribution of the readings and r is the score reliability.

S.E. $(meas.)_1$ using S_1 for distribution estimated as a single reading, but computed from sum of two readings.

S.E. $(meas.)_2$ using S_2 for distribution of assigned scores.

S.E. $(meas.)_p$ standard error of measure reported by publisher. (4:30).

The effect on variability when scores are summed and when they are averaged is clearly evident in the above Table. It was also found that for the assigned or reported score for the Essay Test, the standard error of measure for the Clover Bar students is smaller than that reported by the publisher. This indicates that of a number of essays written by any one Clover Bar student a slightly higher proportion would be judged similar to the sample than would the essay's of any one student in the publisher's norming group. The standard errors of measurement for the Clover Bar group, however, are still sufficiently large to indicate that fairly frequently there would be a score difference of one grade found between essays of equal value.

The statistics would suggest that variability of scores of the Clover Bar group is not as great as that reported by the publisher since the calculation of both standard errors of measurement used the same score reliability coefficient. This reduced variability of Clover Bar students can possibly be attributed either to the homogeneous nature of this group compared to the publisher's norming population, or to the possibility that the potential writing ability of Clover Bar students has not been fully developed in the schools to permit the full play of individual differences, or to both.

Summary

In summary, for the Clover Bar School Division, the technical data provided by the publisher would represent, at best, an underestimate of the quality of the Essay Test. In view of the large number and varied backgrounds of the readers of the Clover Bar Essay Tests, a much poorer inter-reader reliability was expected than was actually found. It stands to the publisher's credit that the scoring procedures, if adhered to, are effective and comparable to the efforts of highly trained readers. The Essay Tests are not without their faults, and many teachers will object to the disregard for the niceties of essay writing, such as, neatness. Yet, within the limitations cited by the publisher, the STEP Essay Tests can make a contribution to the school's educational program by providing a needed standardized instrument for use by the classroom teacher.

BIBLIOGRAPHY

1. Baron, Dennis and Harold W. Bernard. *Evaluation Techniques for Teachers.* New York: McGraw-Hill Book Company, Inc., 1958, p. 206.
2. Bradfield, James M. and H. Stewart Moredock. *Measurement and Evaluation in Education.* New York: The Macmillan Company, 1957, pp. 73-83.
3. College Entrance Examination Board, 425 West 117th Street, New York, 27, N.Y.
4. Cooperative Sequential Tests of Educational Progress. *Handbook for Essay Tests*: (Levels 1-4). Princeton, New Jersey: Cooperative Test Division, Educational Testing Service, 1957.

5. Diederich, Paul B. *The Improvement of Essay Examinations.* A paper delivered at the College Testing Conference, Mount Union College, Alliance, Ohio, February 15, 1957.
6. Educational Testing Service, 20 Nassau Street, Princeton, New Jersey, or 4640 Hollywood Boulevard, Los Angeles 27, California.
7. Ross, C. C. and Julian C. Stanley. *Measurement in Today's Schools.* New York: Prentice-Hall, Inc., 1954, pp. 192-206.
8. Stalnaker, John M. "The Essay Type of Examination." E. F. Lindquist, editor. *Educational Measurement.* Washington, D.C.: American Council on Education, 1951, pp. 495-530.
9. Thorndike, Robert L. and Elizabeth Hagen. *Measurement and Evaluation in Psychology and Education.* New York: John Wiley & Sons. Inc., pp. 35-47.
10. Torgerson, Theodore L. and Georgia Sach Adams. *Measurement and Evaluation for the Elementary-School Teacher.* New York: The Dryden Press, 1954, pp. 227-231.

BOOK REVIEW

The Teaching of Arithmetic. J. V. McIntyre, B.A., B.Paed., Teacher's College, Toronto. J. M. Dent and Sons (Canada) Ltd. 1957, pp. 244, $3.00.

This text is one of a series, the third in publication, entitled Canadian Teachers' Professional Library, under the general editorship of Henry Bowers, principal of Stratford Teachers' College. In the Introduction to this volume Dr. Bowers has approving words for sane methodologies and sane methodologists. Mr. McIntyre is obviously one of the latter. The book may be recommended for use by any teacher in an elementary classroom. It is well organized, well written, and comprehensive.

The chapter headings cover all the topics which one would expect to find in an arithmetic methods book. Some of the subjects are: addition and subtraction, our number system, multiplication, and fractions. The final section is devoted to diagnostic and remedial teaching. "In arithmetic teaching, symptoms of delayed responses, finger counting, poor work habits and gross inaccuracies indicate the necessity for a careful study of the disability and the application of remedial measures." The detailed manner with which errors are analyzed is revealed by the outline accompanying addition:

(a) mastery of basic combinations and extension tables.

(b) addition of single-digit numbers.

(c) addition of two-digit and three-digit numbers.

(d) addition with carrying.

(e) development throughout of accuracy and reasonable speed.

This outline is typical. Teachers are frequently reminded of the sequential nature of arithmetic.

Some general rules are provided:

1. Be sure that you have a clear picture of the sequence of development pupils have followed in each of the basic operations and topics.
2. Follow a scheme that will provide for adequate review of the ideas and skills already learned, the addition of new elements of understanding, and a general improvement in both.
3. When a serious weakness becomes apparent in the work of one or several pupils, locate the area of its occurrence, and apply remedial measures.
4. Always follow remedial teaching with a further test to ensure that your teaching has been effective.

The busy teacher should welcome the clear and concise explanations given in this book.

W. E. Hodgson